Effects of Exposure Imprecision on Estimation of the Benchmark Dose

Risk Analysis (in press)

Effects of Exposure Imprecision on Estimation of the Benchmark Dose

Esben Budtz-Jørgensen[1,2,3], Niels Keiding[1], Philippe Grandjean[2,4]

[1] *Department of Biostatistics, University of Copenhagen*
Blegdamsvej 3, DK-2200 Copenhagen N, Denmark.
[2] *Institute of Public Health, University of Southern Denmark*
Winsløwparken 17, DK-5000 Odense C, Denmark.
[3] *Department of Biostatistics, Harvard School of Public Health,*
Boston, MA 02115, USA
[4] *Department of Environmental Health, Harvard School of Public Health,*
Boston, MA 02115, USA

SUMMARY: In regression analysis failure to adjust for imprecision in the exposure variable is likely to lead to underestimation of the exposure effect. However, the consequences of exposure error for determination of safe doses of toxic substances have so far not received much attention. The benchmark approach is one of the most widely used methods for development of exposure limits. An important advantage of this approach is that it can be applied to observational data. However, in this type of data, exposure markers are seldom measured without error. It is shown that, if the exposure error is ignored, then the benchmark approach produces results that are biased toward higher and less protective levels. It is therefore important to take exposure measurement error into account when calculating benchmark doses. Methods that allow this adjustment are described and illustrated in data from an epidemiological study on the health effects of prenatal mercury exposure.

KEY WORDS: Environmental epidemiology, exposure measurement error, effect of prenatal mercury exposure, exposure standards, benchmark dose.

1 INTRODUCTION

In environmental epidemiology, it is well documented that non-differential measurement error in the exposure variable generally results in an underestimation of the risk.[1,2] As a partial solution to this problem, exposure biomarkers are increasingly used to obtain an estimate of a subject's intake of a particular pollutant under study, and chemical laboratories therefore aim at providing highly reliable results of the biomarker analyses. However, even if laboratory quality data show that the imprecision is very low (e.g., below 5%), this estimate may not reflect the total imprecision, which may also include temporal variability, physiological changes, and individual differences. We recently carried out calculations that showed total imprecisions of 30-50% of two exposure biomarkers[3,4,5,6] that are otherwise considered valid. This degree of exposure misclassification suggests that the impact of imprecisions should be assessed with regard to relevant derived values, such as the benchmark dose.

The so-called benchmark approach is currently one of the most widely used methods in deriving environmental exposure standards. As proposed by Crump[7] the benchmark dose (BMD) is the dose of a substance that increases risk of an abnormal response by a benchmark response (BMR), i.e., from P_0 in unexposed subjects to P_0+BMR at the BMD. The exposure limit or reference dose is then obtained by dividing the lower 95% confidence limit of the BMD, the so-called BMDL, by an uncertainty factor. The present paper shows that the naive analysis that ignores measurement error leads to an overestimation of the BMDL. As expected this overestimation depends on the extent of the measurement error but other study-specific factors, such as the width of the exposure interval, also play a role. Accordingly, application of benchmark results for estimating exposure limits requires that proper account is taken of the overall impact of exposure misclassification. The main part of this paper will describe methods that adjust BMDL calculations for exposure errors.

The consequences of the exposure error in benchmark calculations are illustrated in an epidemiological cohort study conducted in the Faroe Islands to investigate the effect of prenatal methylmercury exposure on childhood cognitive performance. The findings of the Faroe studies have been used to calculate benchmark dose levels that subsequently served as basis for exposure limits.[8] In epidemiological studies of methylmercury toxicity, the hair-mercury concentration has been widely applied as a marker of recent exposures.[9] However, mercury vapor may cause contamination of the hair, while permanent hair treatment may substantially reduce the mercury concentration.[9] Thus, the cord-blood mercury concentration was applied in the Faroes study as the most appropriate marker of prenatal methylmercury exposure.[10] Using factor analysis and structural equation analysis, the cord-blood mercury measurements were recently found to be significantly more accurate than hair-mercury levels.[3,6] We included both exposure variables in the calculations to determine the impact of two different imprecision levels on the benchmark dose results.

2 THE FAROESE MERCURY STUDY

Methylmercury is a common contaminant in seafood and freshwater fish. While adverse effects have been unequivocally demonstrated in poisoning incidents in Japan and in Iraq, the implications of lower-level exposures in fish eating populations have been controversial.[8]

On the Faroe Islands the population is exposed to increased levels of mercury mainly through consumption of contaminated pilot whale meat. To investigate the health effects of this exposure a birth cohort consisting of 1022 children was generated in 1986 and 1987. For each child, information about the prenatal mercury exposure was obtained by measuring mercury concentrations in the cord blood and maternal hair. Then, at 7 years, the children underwent a detailed neuropsychological examination consisting of a number of tests assessing different domains of brain function. Using the cord blood concentration as the exposure variable, ordinary regression analysis showed statistically significant mercury effects for 8 out of the 20 neuropsychological test scores.[10]

In 2000, an expert committee appointed by the National Academy of Sciences (NAS) evaluated the appropriateness of the reference dose for mercury determined by the U.S. Environmental Protection Agency.[8] This committee identified the Faroese study as the critical study for calculations of exposure limits for mercury. Furthermore, the committee recommended that the exposure limit should be calculated based on the benchmark approach, which is described in the next section.

3 THE BENCHMARK APPROACH

The benchmark concept was first developed for standardized experimental dichotomous (normal/abnormal) responses.[7] The benchmark dose (BMD) is the dose that increases the risk of an abnormal response from P_0 in unexposed subjects to P_0+BMR for subjects at the BMD. The excess risk (BMR) must be specified and regulatory agencies often use 5%. Everything else being equal, a lower BMR will result in a lower BMD. To take the statistical uncertainty into account the exposure limit calculation is based on the 95% lower (one-sided) confidence limit of the BMD the so-called BMDL.

Here, attention is restricted to observational continuous response data, thus considering the regression model

$$Y(X, Z) = \beta_0 + \beta_x g(X) + \beta_z^t Z + \epsilon, \tag{1}$$

where Y is the response, X is the exposure, Z is a column of confounders, and the

residual ϵ follows a normal distribution with mean 0 and variance σ^2 $\{\epsilon \sim N(0, \sigma^2)\}$. The dose-response function g is assumed to be known and increasing with $g(0) = 0$. Large responses are assumed to be disadvantageous.

Also for continuous data can the BMD be defined as the dose that increases the risk of an abnormal response from P_0 to $P_0 + \text{BMR}$. Thus, the level t_0 of an abnormal response must be specified. Further, the presence of the confounders complicates the definition of the BMD. It is easily seen that if the abnormal response level (t_0) is defined independent of the confounders, then the BMD will depend on the confounders, which is not practical. This problem can be solved by letting t_0 depend on the confounder values such that all unexposed subjects have the same pre-specified risk P_0 of an abnormal response, ($P\{Y(0, Z) > t_0\} = P_0$ for all Z). With this definition, independently of the confounders the BMD is given by

$$\text{BMD} = \begin{cases} \infty & \text{if } \beta_x \leq 0 \\ g^{-1}(\frac{\Omega\sigma}{\beta_x}) & \text{if } \beta_x > 0 \end{cases} \tag{2}$$

where $\Omega = \Phi^{-1}(1 - P_0) - \Phi^{-1}(1 - P_0 - \text{BMR})$. Often, P_0 is fixed at 5% while the BMR is either 5% or 10%.[8] Note that if $\beta_x \leq 0$ an increasing dose will not increase the expected response and the BMD is infinite.

In a given data set, the BMD is estimated the by substituting the unknown parameters in Equation (2) by their estimates. Furthermore, by exploiting that $\widehat{\beta}/\widehat{\sigma}$ follows a scaled non-central t-distribution, Budtz-Jørgensen *et al.*[11] derived the following expression for the BMDL

$$\text{BMDL} = \begin{cases} \infty & \text{if } \quad t = \widehat{\beta}_x/\widehat{\text{SE}}(\widehat{\beta}_x) \leq -u_{95} \\ g^{-1}\{\frac{\Omega\widehat{\sigma}(1-u_{95}^2/2df)}{\widehat{\beta}_x + u_{95}\widehat{\text{SE}}(\widehat{\beta}_x)\sqrt{1+(t^2-u_{95}^2)/2df}}\} & \text{if } \quad t = \widehat{\beta}_x/\widehat{\text{SE}}(\widehat{\beta_x}) > -u_{95} \end{cases} \tag{3}$$

where $\widehat{\text{SE}}(\widehat{\beta}_x)$ is the estimated standard error of $\widehat{\beta}_x$, df is the number of degrees of freedom, and $u_{95} \approx 1.645$ is the 95'th percentile in the standard normal distribution.

4 ADJUSTMENT FOR EXPOSURE MEASUREMENT ERROR IN BENCHMARK CALCULATIONS

The results of the previous section were derived based on the usual assumption that all independent variables are measured without error. However, this assumption is

not realistic for environmental exposure variables which are typically subject to ordinary laboratory measurement error as well as to biological and other (e.g., temporal) variations. Thus, instead of the true (causative) exposure variable X the investigator is left with an error prone measure W. It is well known that if the measurement error is ignored and X is naively replaced by W in the regression analysis, then the exposure effect is often underestimated.[1,2] A less well known consequence of exposure error is that the residual variance in the dose-response relation is overestimated.[2] Based on Equations (2) and (3) Budtz-Jørgensen *et al.*[11] showed that both BMD and BMDL decrease as a function of the exposure effect and increase as a function of the residual variance. Thus, failure to adjust for measurement error will lead to benchmark results that are biased toward less protective exposure levels.

This section describes how exposure measurement error can be taken into account in the benchmark calculations. There are two main approaches to measurement error correction. Structural methods exploit assumptions about the distribution of the true exposures (X) given the confounders (Z).[12] In functional methods, the true exposures are treated as unknown parameters and distributional assumptions are not needed.[12] Here attention is restricted to structural methods which allow the unknown parameters to be estimated using the maximum likelihood principle. Likewise, we shall consider an extended version of the classic additive error model. Here it is assumed that after transformation with a known monotone function (h) the exposure measurement is given as a sum of the true transformed dose and an independent error, i.e., $h(W) = h(X) + U$, where the error term U is assumed to follow a normal distribution with mean 0 and known variance. The measurement error is said to be non-differential if Y and W are independent given X and Z, i.e., the measured exposure contains no information about the response beyond the information contained in the true exposure.

Assuming independence between subjects and non-differential measurement error, the likelihood function can be written as a product of integrals with respect to the unobserved true exposures

$$L(Y, W, Z, \theta, \gamma) = \prod_{i=1}^{n} f(y_i|w_i, z_i, \theta, \gamma) = \prod_{i=1}^{n} \int f(y_i|x_i, z_i, \theta) f(x_i|w_i, z_i, \gamma)\, dx_i \qquad (4)$$

In the latter expression, the first component $\{f(y_i|x_i, z_i, \theta)\}$ is the dose-response model of interest. The second component is specified by noting that $f(x_i|w_i, z_i) = Cf(w_i|x_i, z_i)f(x_i|z_i)$, where C is a constant not depending on x_i. Thus, in addition to the dose-response model and the error model $\{f(w_i|x_i, z_i)\}$, a model is needed for the distribution of the true exposures given the confounders $\{f(x_i|z_i)\}$. Here we shall assume that $h(X)|Z \sim N(\psi_0 + \psi_z^t Z, \mathrm{var}\{h(X)|Z\})$. This means that the distribution $h(X)|W, Z$ is also normal and therefore the likelihood function can be determined by integration with respect to a normal distribution.

For clarification, it is noted that we are now considering three different exposure scales. The exposure standard is to be placed on the scale of X, $g(X)$ is linearly related to the response while $h(X)$ is normally distributed given the confounders and has additive measurement error. As we shall see, the complexity of the measurement error adjustment depends on the relation between g and h.

4.1 Adjustment in linear models g = h

In this section, measurement error adjusted BMDs and BMDLs are calculated in the special case where the same transformation of the exposure can be assumed to follow the additive error model and to affect the response linearly. Thus, the dose-response model is given by the regression: $Y(X,Z) = \beta_0 + \beta_x g(X) + \beta_z^t Z + \epsilon$, $\epsilon \sim N(0,\sigma^2)$, while the additive error model assumes that $g(W) = g(X) + U$. Under the additional assumption that $g(X)|Z \sim N(\psi_0 + \psi_z^t Z, \mathrm{var}\{g(X)|Z\})$, it is straight forward to show that the conditional distribution of the response given observed exposure and the confounders is: $N(\tilde{\beta}_0 + \tilde{\beta}_x g(W) + \tilde{\beta}_z^t Z, \tilde{\sigma}^2)$, where $\tilde{\beta}_x = \lambda\beta_x$, $\tilde{\sigma}^2 = \sigma^2 + \lambda\beta_x^2\mathrm{var}(U)$ and $\lambda = \mathrm{var}\{g(X)|Z\}/[\mathrm{var}\{g(X)|Z\} + \mathrm{var}(U)]$. The so-called reliability ratio (λ) is less than or equal to 1, so if the true exposure X is naively replaced by the measured exposure W, then the exposure effect is underestimated while the residual variance is overestimated. Therefore, the naive BMD-estimator, calculated ignoring exposure errors, will be biased high. A consistent estimator can be obtained by adjusting the naive estimates ($\widehat{\beta}_x$ and $\widehat{\sigma}^2$) of β_x and σ^2

$$\widehat{\mathrm{BMD}} = g^{-1}\left\{\frac{\Omega\sqrt{\widehat{\sigma}^2 - \widehat{\beta}_x^2\mathrm{var}(U)/\lambda}}{\widehat{\beta}_x/\lambda}\right\} \quad \text{if} \quad \widehat{\beta}_x > 0 \tag{5}$$

In case the reliability ratio is not known, its value can be consistently estimated by exploiting that $\lambda = [\mathrm{var}\{g(W)|Z\} - \mathrm{var}(U)]/\mathrm{var}\{g(W)|Z\}$. The conditional variance $\mathrm{var}\{g(W)|Z\}$ is estimated by the residual variance in the regression of $g(W)$ on Z.

For calculation of the BMDL, it is noted that the distribution of the response given the confounders and the *measured* exposure is also given by a linear regression. Therefore, the naive BMDL is a lower 95% confidence limit of $g^{-1}(\Omega\tilde{\sigma}/\tilde{\beta}_x)$. Thus, $P\{\mathrm{BMDL}_{naive} \leq g^{-1}(\Omega\sqrt{\sigma^2 + \lambda\beta_x^2\mathrm{var}(U)}/\lambda\beta_x)\} \approx 95\%$. Exploiting that the function g is increasing and that $g(\mathrm{BMDL}_{naive}) > 0$ we get

$$P\{g(\mathrm{BMDL}_{naive})^2 \leq g(\mathrm{BMD})^2/\lambda^2 + \Omega^2\mathrm{var}(U)/\lambda\} \approx 95\% \tag{6}$$

and therefore

$$P\{\lambda^2 g(\mathrm{BMDL}_{naive})^2 - \Omega^2 \mathrm{var}(U)\lambda \leq g(\mathrm{BMD})^2\} \approx 95\%. \quad (7)$$

Because $g(\mathrm{BMD})^2 > 0$ the probability is unchanged if negative values of the random variable $\lambda^2 g(\mathrm{BMDL}_{naive})^2 - \Omega^2 \mathrm{var}(U)$ is replaced by zero. Finally, a lower 95%-confidence limit of the BMD can be obtained after transformation by the square root and then by g^{-1}, yielding $P[g^{-1}\{\sqrt{\lambda^2 g(\mathrm{BMDL}_{naive})^2 - \Omega^2 \mathrm{var}(U)\lambda}\} \leq \mathrm{BMD}] \approx 95\%$. Consequently,

$$\mathrm{BMDL} = g^{-1}\{\,\lambda\sqrt{g(\mathrm{BMDL}_{naive})^2 - \Omega^2 \mathrm{var}(U)/\lambda}\,\} \quad (8)$$

is a lower 95% confidence limit for BMD. Here BMDL_{naive} denotes any confidence limit estimate, which would have had the correct coverage probability had the exposure variable been error free.

From Equation (8), it is seen that the naive BMDL is higher than the adjusted value. Thus, if exposure error is ignored in the benchmark calculation estimated exposure limits will be too high. The larger the exposure imprecision the stronger is this overestimation. On the other hand, an increased variance in the true exposures (given the confounders) will limit the bias.

It is should be noted that these calculations ignore the estimation uncertainty associated with the reliability ratio. The coverage probability of the adjusted BMDL (8) may therefore differ slightly from the nominal value. The estimation uncertainty in λ may be taken into account by applying the bootstrap-based algorithm described the next section.

4.2 Adjustment in nonlinear models g ≠ h

In the general case, $g(X)$ is linearly related to the response $\{Y = \beta_0 + \beta_x g(X) + \beta_z^t Z + \epsilon\}$, while a different transformation $h(X)$ is required for normality and additive measurement error $[h(X)|Z \sim N(\psi_0 + \psi_z^t Z, \mathrm{var}\{h(X)|Z\})$ and $h(W) = h(X) + U]$. Under these weaker assumptions, more sophisticated statistical methods are required for measurement error adjustment. As shown in Equation (4), the likelihood function is given as a product of integrals with respect to the conditional distribution of the unobserved exposures given the measured exposure and the confounders

$$\prod_{i=1}^{n} \int \phi([y_i - \beta_0 - \beta_x g\{h^{-1}(s_i)\} - \beta_z^t z_i]/\sigma)\, \phi[\{s_i - \alpha_0 - \alpha_w h(w_i) - \alpha_z^t z_i\}/\delta]/\sigma\delta\, ds_i \quad (9)$$

where $s = h(x)$, ϕ is the density of the standard normal distribution and $\delta^2 = \mathrm{var}\{h(X)|h(W), Z\}$. Since the composite function gh^{-1} may be non-linear, a general

closed form expression for the likelihood cannot be obtained. An estimation method based on numerical integration is therefore described.

The parameters are estimated in a two step procedure. In the first step, the parameters $(\alpha_0, \alpha_w, \alpha_z, \delta^2)$ of the distribution of $h(X)$ given $h(W)$ and Z are estimated. This is done based on the fact that these parameters are functions of the parameters in the distribution $h(X)|Z$ and the reliability ratio: $(\alpha_0, \alpha_z) = (1-\lambda)(\psi_0, \psi_z)$, $\alpha_w = \lambda$ and $\delta^2 = \text{var}\{h(X)|h(W), Z\} = (1-\lambda)\text{var}\{h(X)|Z\}$. The reliability ratio and ψ-parameters are estimated by regressing $h(W)$ on Z. Estimates of $(\alpha_0, \alpha_w, \alpha_z, \delta^2)$ are then inserted in the likelihood function (9). In the second step, the likelihood function (9) is maximized as a function of $(\beta_0, \beta_x, \beta_z, \sigma^2)$ using the SAS-procedure NLMIXED. Finally, the BMD is estimated by inserting the measurement error adjusted estimates of β_x and σ in the Equation (2).

The lower confidence limit of the BMD, the BMDL, is approximated by the 5th percentile in the estimated distribution of $\widehat{\text{BMD}}$. This percentile is determined using the parametric bootstrap.[13] For each subject, a new exposure and response value $(\tilde{W}, \tilde{Y})$ are generated by sampling $\tilde{W}$ and $\tilde{Y}$ from the conditional distribution of W and Y given Z estimated from original data. This can be done by first sampling $\tilde{X}$ from the distribution $f[x|z, \widehat{\psi}, \widehat{\text{var}}\{h(X)|Z\}]$ whereupon $\tilde{W}$ is sampled based on the additive error model $f\{w|\tilde{x}, \text{var}(U)\}$. Finally, $\tilde{Y}$ is sampled from the estimated dose-response model $f(y|\tilde{x}, z, \widehat{\theta})$. In the new data set, the BMD is estimated using the procedure described above. In the general case where g and h can be different, this requires application of NLMIXED for each bootstrap sample, while the BMD can be estimated using the simpler approach of Section (4.1) in case g and h are equal. The whole procedure is repeated a large number of times and the BMDL is calculated as the 5th percentile in the empirical distribution of the BMD-estimates. Note that, because the whole estimation process is repeated in each bootstrap sample, this procedure also takes into account any estimation uncertainty associated with the parameters determined in the first steep $(\alpha_0, \alpha_w, \alpha_z, \delta^2)$.

As an alternative to bootstrap sampling, an approximation may be obtained by inserting measurement error adjusted estimates of $\beta_x, \sigma^2, \text{var}(\widehat{\beta}_x)$ in the BMDL expression (3). This expression was derived based on the fact that, in case of zero exposure imprecision, the ordinary regression estimates of β_x and σ^2 are unbiased, independent and follow a normal and a χ^2-distribution, respectively. If the adjusted estimates obtained from NLMIXED have approximately the same properties, then this statistic will have a coverage probability close to the nominal value.

5 ADJUSTED BENCHMARK CALCULATIONS FOR PRENATAL METHYLMERCURY EXPOSURE

Having chosen the Faroese study, the NAS expert committee decided that the exposure limit should be calculated based on the outcome with the strongest mercury effect, the so-called Boston Naming Test. Both P_0 and the BMR were fixed at 5%, which means that exposure to the corresponding BMD will double the risk of an abnormal response (from 5% to 10%). For these data, BMD results were quite sensitive to the choice of dose-response model;[11] but the NAS committee did not find a square root and a logarithmic model biologically plausible and instead recommended a linear model for the relation between mercury concentrations and the neuropsychological outcome. The original calculation yielded a BMDL of 58 μg/l (Table I), which was used as the so-called point of departure in the identification of the reference dose.[8] However, this calculation did not take exposure measurement error into account and the BMDL is likely to have been overestimated.

Application of the adjustment method described above requires information about the size of the measurement error in the mercury biomarkers. In environmental epidemiology, the total exposure imprecision is a sum of two different types of error: laboratory measurement imprecision and biological variation. In the case of prenatal mercury exposure, the second (pre-laboratory) error component is a consequence of the fact that the mercury concentration in the fetal circulation is not constant over time, but varies according to maternal mercury intake. It may also include individual differences in the distribution of mercury in the body and across the placenta.

Information about the size of laboratory error is typically available from quality control data, but the total error is generally unknown. In this study, total imprecisions have been estimated from the joint distribution of multiple exposure markers using factor analysis and structural equation models.[3,4,5,6] After a logarithmic transformation, mercury concentrations in cord blood ($B\text{-}Hg$) and maternal hair ($H\text{-}Hg$) are approximately linearly related with a homogeneous scattering.[3,4] A simple factor analysis model may therefore be appropriate for the joint distribution of the exposure variables. This model assumes that the i'th child's (transformed) biomarker values depend linearly on the true (transformed) exposure $\{\log(X_i)\}$ and a random error

$$\begin{aligned} \log(B\text{-}Hg)_i &= \log(X_i) + U_{B\text{-}Hg,i} \\ \log(H\text{-}Hg)_i &= \nu_{H\text{-}Hg} + \gamma_{H\text{-}Hg} \cdot \log(X_i) + U_{H\text{-}Hg,i} \end{aligned} \tag{10}$$

In this model, the true exposure is expressed on the scale of the cord blood concentrations in the sense that a one unit increase in $\log(X_i)$ on average leads to a one unit increase in $\log(B\text{-}Hg)_i$. The factor loading $\gamma_{H\text{-}Hg}$ allows for the fact that a one unit increase in $\log(B\text{-}Hg)_i$ may not correspond to a one unit increase in $\log(H\text{-}Hg)_i$.

Similarly the intercept ($\nu_{H\text{-}Hg}$) allows for differences in origins. The parameters of main interest here are the variances of the error terms $U_{B\text{-}Hg,i}$ and $U_{H\text{-}Hg,i}$.

The parameters in the factor analysis model cannot be estimated based on data from only two exposure variables. This problem was solved by exploiting information about maternal whale consumption, the main source of prenatal mercury exposure. Thus, the mother's average number of pilot whale dinners per month during pregnancy ($Whale$) was included as a predictor of the true exposure

$$\log(X_i) \quad = \quad \mu + \kappa \cdot \log(Whale_i + 1) + \zeta_i \tag{11}$$

Here ζ_i is a normally distributed residual term with mean zero. Figure I gives a graphical illustration of this model.

Figure I here

This analysis showed that the cord blood concentration gives the most precise reflection of the true exposure. This result is in agreement with prior expectations and with the empirical finding that the cord blood concentration was a stronger predictor of childhood coginitive deficits.[14] The estimated error variance in the log-transformed cord blood concentration was 0.101 corresponding to a reliability ratio of 0.865 and an error CV of about 32% in the untransformed concentration. For the maternal hair concentration, the reliability ratio was 0.681. These results were confirmed in more sophisticated structural equation modeling, which allowed inclusion of information from the neuropsychological outcomes in the calibration of the exposures. A parsimonious model with a close fit to the data was developed by viewing also the child's test scores as manifestations of underlying variables.[5,6] The association between exposure and outcome variables was described by assuming that the latent exposure affected the latent neuropsychological functions.

Table I shows benchmark results for the cord blood and the maternal hair concentration before and after adjustment for measurement errors. With a logarithmic error model and a linear dose-response model, a non-linear model (section 4.2) is required to perform the adjustment in this case. Adjusted BMDLs are calculated in two different ways: using the parametric bootstrap and by inserting adjusted estimates in Equation (3).

It is seen that measurement error adjustment leads to clear reductions in the benchmark results. The cord blood BMDL is reduced by almost 25% while the BMDL of the more uncertain hair concentration is reduced by approximately 40%. Furthermore, BMDLs based on Equation (3) are seen to be in close agreement with the results obtained through bootstrap simulation.

Coverage probabilities of the confidence limits can be explored in the (parametric) bootstrap samples used for BMDL calculation. In each simulated data set, BMDL values are calculated using each of the methods described above, except the bootstrap

method itself. Since the BMD is known under the simulations, coverage probabilities can be estimated as the relative frequency of data sets, in which the BMDL is lower than the BMD. Not surprisingly, the coverage probability of unadjusted BMDLs is seen to be clearly below the nominal value. In contrast, for both exposure variables the adjusted BMDL based on Equation (3) has coverage probabilities close to 95%. Estimation of the performance of the parametric bootstrap would be very computer intensive as this would require a new bootstrap study for each of the 15,000 data sets. However, the good performance of the method based on Equation (3) and the close agreement between this method and the parametric bootstrap in the original data, indicate that the parametric bootstrap BMDL also has a coverage probability close to 95%.

Table I here

6 DISCUSSION

The purpose of the benchmark approach is to allow for statistical uncertainty when determining a well defined level-of-departure. This aim is achieved in the sense that small studies with fewer subjects will tend to produce lower BMDLs.[7] However, studies in environmental epidemiology typically involve many sources of uncertainty, in addition to a limited sample size. The present study has focused on exposure imprecision. The results document that if a non-differential measurement error is ignored, then the BMD and the BMDL are overestimated. Thus, application of a less precise biomarker will lead to a higher and less protective exposure limit. This effect is in accordance with the known bias toward the null in regression coefficients.[1,2] However, this bias is counter to the precautionary principle, which would require that weaker knowledge must lead to more stringent standards.[15,16]

Few exposure variables are error-free in observational studies. The present results suggest that efforts would be worthwhile to base benchmark calculations on the exposure biomarkers of the best validity. Even in that case, estimations of the impact of inevitable imprecision should be conducted. Unfortunately, detailed information on the exact imprecision of exposure parameters is typically not available in epidemiological studies, and unbiased benchmark results may therefore not be obtainable. In such cases, alternative methods may be considered. Hockey stick models assume no effect of the exposure below an unknown threshold. In the presence of an exposure imprecision the threshold is typically underestimated.[17] Thus, contrary to the benchmark approach, the naive analysis that ignores the measurement error will produce lower hockey stick thresholds and thereby more protective standards.

In the case of methylmercury exposure, this paper explored the extent to which estimated exposure errors affect BMDLs already published and used in reference dose calculation.[8,11] Based on measurement imprecisions that had been estimated

from the study results themselves, it was shown that the (adjusted) BMDL was overestimated by about 30% and more than 60% for the two exposure biomarkers. The adjustments needed are rather limited in this case, because of the wide exposure interval. Thus, these calculations may represent only the lower range of adjustments that may typically be required. The relative magnitude of the adjustments may be appreciated when compared with the uncertainty factors that are usually applied when estimating exposure limits.

When BMDLs are used for calculation of a dose that is considered safe to ingest, an uncertainty factor is applied. In the case of methylmercury, an uncertainty factor of 10 was chosen, so that the exposure limit was set at one tenth of the BMDL.[8] The uncertainty factor must take into account the degree of individual variation in the toxicokinetic and toxicodynamic aspects of methylmercury toxicity, each of which is often assigned a factor of 3.2 (i.e., the square root of 10).[18] The toxicokinetic factor allows individual differences in the relation between the external (ingested) dose and the chemical concentration at the site of action while the toxicodynamic factor allows for inter-individual differences in the response, in relation to the same concentration at the site of action.[18] Thus, the toxicokinetic part of the uncertainty includes variations in the uptake and transfer of methylmercury in the body. It therefore also includes physiological variations that affect the hair and blood mercury concentrations. Several studies have addressed this issue on the basis of observational studies of the hair/blood concentration ratio.[19] However, by basing the uncertainty factor on observational studies, it also includes a certain amount of measurement error. A more complete physiologically-based toxicokinetic model would be more appropriate,[19] but the lack of information on specific processes makes the use of default assumptions necessary, thereby limiting the validity of the calculations. Accordingly, measurement uncertainty is also involved in the decision on uncertainty factors, but it refers to the individual studies used to determine the toxicokinetic variability. Adjustment of BMDLs for measurement error in the dose-response studies would therefore not seem to cause any double-counting of the uncertainty. Thus, uncertainty factors probably need not be changed because of the present considerations. The dependence of BMDLs on exposure imprecision is therefore likely to result in overestimations of exposure limits, the degree of which depends on the characteristics of the study used for the BMDL calculation.

REFERENCES

1. Fuller, W.A. (1987). *Measurement Error Models.* Wiley.

2. Carroll, R.J. (1998). Measurement error in epidemiologic studies. In Armitage P, Colton T (Eds.), *Encyclopedia of biostatistics* (pp. 2491-2519). Chichester: John Wiley & Sons.

3. Budtz-Jørgensen, E., Keiding, N., Grandjean, P., Weihe, P. & White, R.F.

(2003). Consequences of exposure measurement error in environmental epidemiology. *Statistics in Medicine, 22*, 3089-3100.

4. Budtz-Jørgensen, E., Grandjean, P., Jørgensen, P.J., Weihe, P. & Keiding, N. Association between mercury concentrations in blood and hair in methylmecury-exposed subjects at different ages. (Accepted for publication in *Environmental Research.*)

5. Budtz-Jørgensen, E., Keiding, N., Grandjean, P., Weihe, P. & White, R.F. (2003). Statistical methods for the evaluation of health effects of prenatal mercury exposure. *Environmetrics, 13*, 105-120.

6. Budtz-Jørgensen, E., Keiding, N., Grandjean, P. & Weihe. P. (2002). Estimation of health effects of prenatal mercury exposure using structural equation models. *Environmental Health, 1*, 2.

7. Crump, K. (1984). A new method for determining allowable daily intakes. *Fundamental and Applied Toxicology, 4*, 854-871.

8. National Academy of Sciences (NAS) (2000). *Toxicological effects of methylmercury.* National Academy Press.

9. Grandjean, P., Jørgensen, P.J. & Weihe, P. (2002). Validity of mercury exposure biomarkers. In: Wilson SH, Suk WA, (Eds.), *Biomarkers of Environmentally Associated Disease* (pp. 235-47). Boca Raton, FL, CRC Press/Lewis Publishers.

10. Grandjean, P., Weihe, P., White, R.F., Debes, F., Araki, S., Yokoyama, K., Murata, K., Sørensen, N., Dahl, R. & Jørgensen, P.J. (1997). Cognitive deficit in 7-year-old children with prenatal exposure to methylmercury. *Neurotoxicology and Teratology, 19*, 417-428.

11. Budtz-Jørgensen, E., Keiding, N., Grandjean, P. (2001). Benchmark dose calculation from epidemiological data. *Biometrics, 57*, 698-706.

12. Carroll, R.J., Ruppert, D. & Stefanski, L.A. (1995). *Measurement error in nonlinear models.* Chapman & Hall.

13. Efron, B. & Tibshirani, R.J. (1993). *An introduction to the bootstrap.* London: Chapman & Hall.

14. Grandjean, P., Budtz-Jørgensen, E., White, R.F., Jørgensen, P.J., Weihe, P., Debes, F. & Keiding, N. (1999). Methylmercury exposure biomarkers as indicators of neurotoxicity in children aged 7 years. *American Journal of Epidemiology, 150*, 301-305.

15. Barnett, V. & O'Hagan, A. (1997). *Setting environmental standards: The statistical approach to handling uncertainty and variation.* London: Chapman & Hal.

16. Keiding, N. & Budtz-Jørgensen, E. (2003). The precautionary principle and statistical approaches to uncertainty. *European Journal of Oncology Library, 2*, 185-191.

17. Kuchenhoff, H. & Carroll, R.J. (1997). Segmented regression with errors in predictors: Semi-parametric and parametric methods. *Statistics in Medicine, 16*, 169-18.

18. Renwick, A.G. & Lazarus, N.R. (1998). Human variability and noncancer risk assessment–an analysis of the default uncertainty factor. *Regulatory toxicology and pharmacology, 27*, 3-20.

19. Stern, A.H., Clewell, H.J. & Swartout, J. (2002). An objective uncertainty factor adjustment for methylmercury pharmacokinetic variability. *Human and Ecological Risk Assessment, 8*, 885-894.

Table

Table I. BMD and BMDL calculated using respectively the cord blood mercury concentration and the maternal hair concentration as the exposure variable. The first line shows the naive results ignoring exposure error. Then, adjusted results based on NLMIXED analysis are given. Adjusted BMDLs are calculated in two different ways: using the 15.000 parametric bootstrap samples and by inserting adjusted estimates in Equation (3). For all BMDLs, except the bootstrap estimate, "cov.prob" indicates the empirical coverage probability in the bootstrap samples.

	Cord blood[a] (μg/l)			Maternal hair[b] (μg/g)		
Method	$\widehat{\text{BMD}}$	BMDL	cov.prob. (%)	$\widehat{\text{BMD}}$	BMDL	cov.prob.(%)
Unadjusted	84.98	57.61	74.74	15.22	10.02	38.46
Parametric bootstrap	65.13	45.50	-	9.51	6.23	-
Equation (3)	65.13	44.08	94.81	9.51	5.99	94.97

SE(cov.prob) $\approx 0.18\%$, $^a\widehat{corr}(\widehat{\beta}_x, \widehat{\sigma}^2) = -0.04$, $^b\widehat{corr}(\widehat{\beta}_x, \widehat{\sigma}^2) = -0.09$

Figur

Figure I. Path diagram illustrating model used in estimation of the measurement error variance. After a logarithmic transformation mercury concentrations in cord blood and maternal hair are considered mainfestations of the underlying true exposure variable. This true exposure is assumed to be affected by the maternal pilot whale meat consumption during pregnancy.